# Module 7: Multilevel Models for Binary Responses

# Stata Practical

*George Leckie* [1]
Centre for Multilevel Modelling

**Pre-requisites**

- Modules 1-6

## Contents



[1] This Stata practical is adapted from the corresponding MLwiN practical: Steele, F. (2008) Module 7: Multilevel Models for Binary Responses. LEMMA VLE, Centre for Multilevel Modelling. Accessed at http://www.cmm.bris.ac.uk/lemma/course/view.php?id=13.





**Most of the sections within this module have online quizzes for you to test your understanding. To find the quizzes:**

From within the LEMMA learning environment

- Go down to the section for **Module 7: Multilevel Models for Binary Responses**
- Click " 7.1 Two-Level Random Intercept Model" to open Lesson 7.1
- Click Q 1 to open the first question

## Introduction to the Bangladesh Demographic and Health Survey 2004 Dataset

You will be analysing data from the Bangladesh Demographic and Health Survey (BDHS),[1] a nationally representative cross-sectional survey of women of reproductive age (13-49 years).

Our response variable is a binary indicator of whether a woman received antenatal care from a medically-trained provider (a doctor, nurse or midwife) at least once before her most recent live birth. To minimise recall errors, the question was asked only about children born within five years of the survey. For this reason, our analysis sample is restricted to women who had a live birth in the five-year period before the survey. Note that if a woman had more than one live birth during the reference period, we consider only the most recent.

These data were analysed in Module 6 using single-level models. In this module, we consider multilevel models to allow for and to explore between-community variance in antenatal care. The data have a two-level hierarchical structure with 5366 women at level 1, nested within 361 communities at level 2. In rural areas a community corresponds to a village, while an urban community is a neighbourhood based on census definitions.

We consider a range of predictors. At level 1, we consider variables such as a woman's age at the time of the birth and education. Level 2 variables include an indicator of whether the region of residence is classified as urban or rural. We will also derive community-level measures by aggregating woman-level variables, for example the proportion of respondents in the community who are in the top quintile of a wealth index.

[1] We thank MEASURE DHS for their permission to make these data available for training purposes. Additional information about the 2004 BDHS and other Demographic and Health Surveys, including details of how to register for a DHS Download Account, is available from www.measuredhs.com.

The file contains the following variables:

| Variable name | Description and codes |
|---|---|
| **comm** | Community identifier |
| **womid** | Woman identifier |
| **antemed** | Received antenatal care at least once from a medically-trained provider, e.g. doctor, nurse or midwife (1 = yes, 0 = no) |
| **bord** | Birth order of child (ranges from 1 to 13) |
| **mage** | Mother's age at the child's birth (in years) |
| **urban** | Type of region of residence at survey (1 = urban, 0 = rural) |
| **meduc** | Mother's level of education at survey (1 = none, 2 = primary, 3 = secondary or higher) |
| **islam** | Mother's religion (1 = Islam, 0 = other) |
| **wealth** | Household wealth index in quintiles (1 = poorest to 5 = richest) |

The dataset also contains a number of extra variables derived from those above (see the practical for Module 6).





# P7.1 Two-Level Random Intercept Model

Load "7.1.dta" into memory and open the do-file for this lesson:

From within the LEMMA Learning Environment
- Go to **Module 7: Multilevel Models for Binary Responses**, and scroll down to ***Stata* Datasets and Do-files**
- Click "7.1.dta" to open the dataset

and use the `describe` command to produce a summary of the dataset:

```
. describe

Contains data from 7.1.dta
  obs:         5,366
 vars:            17                          5 Sep 2009 09:38
 size:       177,078 (99.9% of memory free)
--------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
--------------------------------------------------------------------------------
comm            int    %9.0g                  Community ID
womid           int    %9.0g                  Woman ID
antemed         byte   %9.0g                  Antenatal from qualified medic
bord            byte   %9.0g                  Birth order
mage            byte   %9.0g                  Mother's age at birth
urban           byte   %9.0g                  Type of region of residence
meduc           byte   %9.0g                  Maternal education
islam           byte   %9.0g                  Religion
wealth          byte   %9.0g                  Wealth index (1 = poorest)
magec           float  %9.0g
magecsq         float  %9.0g
meduc2          byte   %8.0g
meduc3          byte   %8.0g
wealth2         byte   %8.0g
wealth3         byte   %8.0g
wealth4         byte   %8.0g
wealth5         byte   %8.0g
--------------------------------------------------------------------------------
Sorted by:
```

## P7.1.1 Specifying and estimating a two-level model

We will begin by fitting a null or empty two-level model, that is a model with only an intercept and community effects.

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + u_{0j}$$

The intercept $\beta_0$ is shared by all communities while the random effect $u_{0j}$ is specific to community *j*. The random effect is assumed to follow a normal distribution with variance $\sigma^2_{u0}$.

Stata's main command for fitting multilevel models for binary response variables is the `xtmelogit` command.[2] The syntax for `xtmelogit` is similar to that for `xtmixed`. To fit the above model using the `xtmelogit` command, we type: `xtmelogit antemed || comm:, variance`.

The binary response variable (**antemed**) follows the command which is then followed by the list of fixed part explanatory variables (excluding the constant as this is included by default[3]). The above model contains only an intercept and so no fixed part explanatory variables are specified. The level 2 random part of the model is specified after two vertical bars `||`. The level 2 identifier (**comm**) is specified first followed by a colon and then the list of random part explanatory variables (again excluding the constant as this is included by default). Finally, the `variance` option reports the variances of the random intercept and any random coefficients included in the model (as opposed to the default of standard deviations).

Issuing the `xtmelogit` command gives the following output:

```
. xtmelogit antemed || comm:, variance

Refining starting values:

Iteration 0:   log likelihood = -3321.6208
Iteration 1:   log likelihood = -3313.2849
Iteration 2:   log likelihood = -3313.2818

Performing gradient-based optimization:

Iteration 0:   log likelihood = -3313.2818
Iteration 1:   log likelihood = -3313.2817
```

[2] Note, two-level random intercept logit models can equally be fitted with the `xtlogit` command; see `help xtlogit`. To fit the equivalent model with the probit link function, see `help xtprobit`. We do not discuss the `xtlogit` or `xtprobit` commands as they cannot be used to fit more complicated multilevel models while `xtmelogit` can. However, we do note that `xtlogit` fits models considerably faster than `xtmelogit` and is therefore recommended for fitting two-level random intercept logit models. See Rabe-Hesketh and Skrondal (2008) for examples of two-level random intercept models fitted with both commands.
[3] Note, the `noconstant` option can be used to omit the constant from the fixed or the random part of the models; see `help xtmelogit`.





```
Mixed-effects logistic regression               Number of obs      =      5366
Group variable: comm                            Number of groups   =       361

                                                Obs per group: min =         3
                                                               avg =      14.9
                                                               max =        25

Integration points =   7                        Wald chi2(0)       =         .
Log likelihood = -3313.2817                     Prob > chi2        =         .

------------------------------------------------------------------------------
     antemed |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .1486212   .0727516     2.04   0.041     .0060307    .2912118
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
comm: Identity               |
                  var(_cons) |   1.502371   .1591921      1.220628    1.849145
------------------------------------------------------------------------------
LR test vs. logistic regression: chibar2(01) =   808.64 Prob>=chibar2 = 0.0000
```

Before interpreting the model, we will discuss the estimation procedure that `xtmelogit` uses. As will be described in C7.7 (and in more detail in the Technical Appendix), there are several estimation procedures available for binary and other categorical response models. However, in Stata, only one procedure is implemented: maximum likelihood estimation using adaptive quadrature. As with the other procedures, this is an approximate method and so it is always important to assess whether the approximation is adequate. By default, `xtmelogit` uses adaptive quadrature with 7 integration points. To check that 7 integration points is adequate, the model can be refitted with a larger number of quadrature points (the `intpoints()` option is used to do this). If the two sets of model parameters are substantially the same, then 7 integration points is adequate. It might also be the case that 7 integration points is more than adequate, in which case the model can be fitted with fewer points.

Table 7.1 gives the parameter estimates which are obtained for the above model when different numbers of integration points are specified: 1, 2, 3, 4, 5, 6, 7 and 15. The percentage difference between each parameter estimate and its most accurate estimate (i.e. when 15 integration points are used) is also reported.[4] The last row of the table reports the time (in seconds) that it takes for the model to converge.[5]

[4] Note that using a higher number of integration points than 15 will lead to more accurate estimates. However, the results in Table 7.1 suggest that little will be gained by doing this.
[5] We used a 64bit 2-core multiprocessor version of Stata 11 on a 2.66Ghz Intel Xeon X7460 running on Windows Server 2008.





*Table 7.1. Estimates for different numbers of integration points reported with the percentage difference between each estimate and that based on 15 integration points*

| **Parameter** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **15** |
|---|---|---|---|---|---|---|---|---|
| $\hat{\beta}_0$ | 0.148 | 0.148 | 0.148 | 0.149 | 0.149 | 0.149 | 0.149 | 0.149 |
| | -0.7% | -0.7% | -0.7% | 0.0% | 0.0% | 0.0% | 0.0% | |
| | | | | | | | | |
| $\hat{\sigma}^2_{u0}$ | 1.464 | 1.464 | 1.483 | 1.501 | 1.500 | 1.502 | 1.502 | 1.503 |
| | -2.6% | -2.6% | -1.3% | -0.1% | -0.2% | -0.1% | -0.1% | |
| | | | | | | | | |
| Log likelihood | -3318 | -3317 | -3314 | -3313 | -3313 | -3313 | -3313 | -3313 |
| | | | | | | | | |
| Seconds | 3.8 | 3.3 | 3.4 | 2.8 | 2.9 | 3.1 | 2.4 | 3.3 |

The table shows that when 1 integration point is used, the constant is 0.7% smaller than when 15 points are used while the between-community variance is 2.6% smaller than its corresponding value. However, increasing the number of integration points to 4 gives an estimate for the variance which is only 0.1% smaller than when 15 integration points are used. All 8 models took a similar length of time to converge. We will therefore continue to use the default setting of 7 integration points in the following random intercept models. However, when we come to specify random slope models (P7.5), we will revisit this issue to check that 7 integration points is still appropriate.

## P7.1.2 Interpretation of the null two-level model

From the model estimates (using 7 integration points), we can say that the log-odds of receiving antenatal care from a medically-trained provider in an 'average' community (one with $u_{0j} = 0$) is estimated as $\hat{\beta}_0 = 0.149$. The intercept for community *j* is $0.149 + u_{0j}$, where the variance of $u_{0j}$ is estimated as $\hat{\sigma}^2_{u0} = 1.502$.

The likelihood ratio statistic for testing the null hypothesis that $\sigma^2_{u0} = 0$ is reported in the final line of the output. The test statistic is 808.64 with a corresponding p-value of less than 0.00005 and so there is strong evidence that the between-community variance is non-zero.[6]

We will now examine estimates of the community effects or residuals, $\hat{u}_{0j}$, obtained from the null model. To calculate the residuals and produce a 'caterpillar plot' with the community effects shown in rank order together with 95% confidence intervals we can use the same commands as we used P5.1.2 for the

[6] Note that the test statistic has a non-standard sampling distribution as the null hypothesis of a zero variance is on the boundary of the parameter space; we do not envisage a negative variance. In this case the correct p-value is half the one obtained from the tables of chi-squared distribution with 1 degree of freedom. For this model, Stata automatically reports the correct p-value.

This document is only the first few pages of the full version.

To see the complete document please go to learning materials and register:

http://www.cmm.bris.ac.uk/lemma

**The course is completely free**. We ask for a few details about yourself for our research purposes only. We will not give any details to any other organisation unless it is with your express permission.